General comments:

I enjoyed reading this elaborated work on disentangling the hydrological and hydraulic controls on streamflow variability in E3SM V2. It is clearly written and organized, however missing some consistency in e.g. naming of experiments, and some figures could use a revision. Those adjustments in the text and figures would make it even more easy to follow the story in the text.

Xu et al. performed several experiments with the E3SM coupled with MOSART. The topic of studying parameter values of critical parameters of a land surface model is crucial, and especially the validation of model output against different observation products is important. It is a valuable contribution to the land surface modelling community.

The case study area of the Pantanal Region is interesting as the streamflow show a shift of about five months in the seasonality compared to precipitation.

Response: We thank the reviewer for the prompt review and constructive comments. We believe your comments are important and very helpful for improving the manuscript. In the revised manuscript, we will modify the experiment names, and update the figures as suggested. Please find our point-to-point response in the following.

I would like some more information about the calibration. I like the simplicity in the random sampling. Are 1000/2000 simulations an appropriate number? Why?

In my world calibration involves a mathematical optimizer, but your approach of doing several experiments with random sampling is an easy and simple approach to get an idea op optimal parameter values. It is not a demand, but maybe you could give some information on how the objective function distribute, and the performance of the “best” solution compared to the default? I would also like to know the parameter values of the “best” calibration. Those numbers would be valuable for other modelers.

Response: The performance of model with calibrated parameter values should increase as the number of calibration simulation increases. According to our previous calibration experience in E3SM (Xu et al., 2022), the three selected parameters in this study are the most sensitive model parameters for simulating runoff process. Randomly selecting 10 values on each parameter lead to a good coverage of the random variable space of the uncertainty parameters. For example, three parameters were selected for calibration in both ELM and MOSART, which leads to $10 \times 10 \times 10 = 1{,}000$ parameter combination in each calibration step. To further investigate if the 1,000 is enough for the calibration, we analyzed the calibrated model performance against the number of simulations used in the calibration. Figure R1a shows that the value of ELM’s objective function (i.e., Step 1 calibration) decreases rapidly as the number of calibration simulations increases to 200 and the decrease in the objective function is much smaller for additional simulations. The initial set of 1,000 simulations was extended to include an additional 500 simulations with random parameters values, but those additional simulations did not further reduce the objective function significantly. This implies that the parameter calibrated with 1,000 simulations is a good approximation to the best parameter (e.g., the parameter calibrated with a much larger number of simulations). Similarly, the performance in step 2 (e.g., MOSART calibration) did not increase after we increase the number of calibration simulations to be larger than ~100 (Figure R1b). The corresponding objective function and correlation coefficients from all the simulations of step 1 and step 2 can be found in the inserts of Figure R1. The two-step calibration significantly improves the model performance as the objection function with default parameter is 20.08 and the correlation coefficient with default simulated streamflow and observed streamflow is 0.2.

We will include the analysis for the number of calibrations in the revised manuscript. We will also add the calibrated parameter values in the supplementary material.

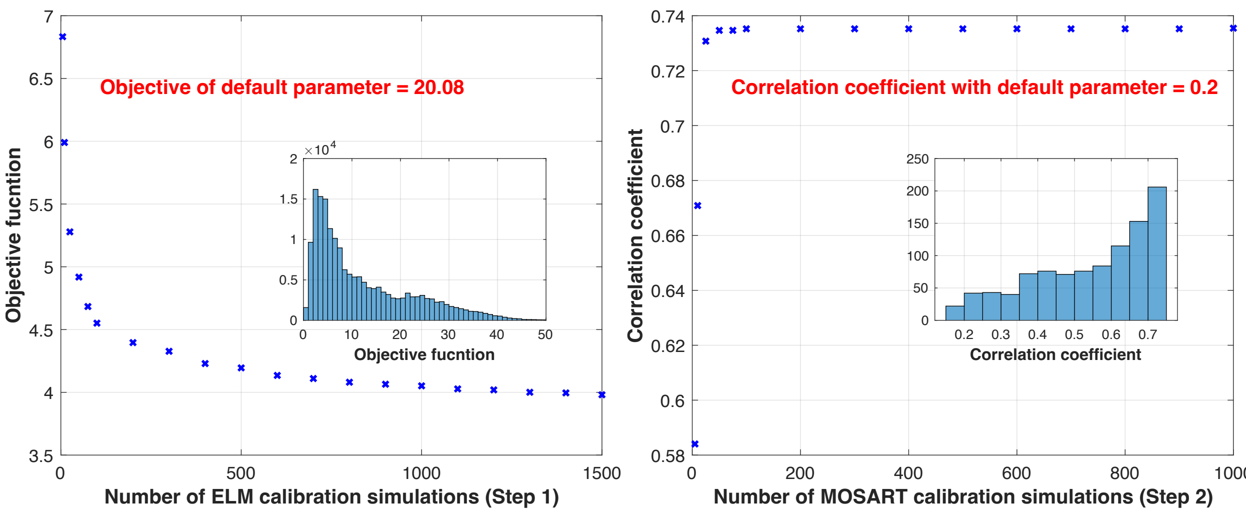

Figure R1. Change of objective function of step 1 calibration (left panel) and correlation between calibrated simulation and observation of step 2 calibration with number of calibration simulations. The histograms in left and right panels illustrates the distribution of the objective function and correlation coefficients from all the simulations during step 1 and step 2 calibration, respectively.

Specific comments

Page 7 line 169: please explain why it is an acceptable assumption.

Response: We made assumption that the annual runoff trend is spatially uniform within the watershed because there no reliable gridded runoff data to derive annual runoff trend at grid level. Although there are several subbasin gauges in the watershed, they have a shorter temporal coverage than the simulation period (i.e., 1979-2009). Therefore, the only way for us to include the trend in the objective function during the calibration is to assume the annual runoff trend is spatially uniform within the watershed. We will add a discussion of this limitation in the revised manuscript.

Figure 4: It is very hard to see difference in the size of the circles. The plot with rSD at xaxis: what is on the yaxis? Probably Manning coef., but this it not obvious. Please improve figure.

Response: Thanks for the suggestion. The reviewer is correct that the plot with rSD has Manning coefficient in the y-axis, which share the ticks with the colorbar. We will update Figure 4 in the revised manuscript to make it clearer.

Page 11 line 238: dosen´t it say some other numbers on figure?

Response: Thanks for catching the typo. We confirm the values reported in the figure is correct, and the value reported in the main text was from previous analysis. We will correct it in the revised manuscript.

Figure 6: I suggest to improve the readability of the figure (only a suggestion, not a need):

- (b): I am confused about what is on the left yaxis
- perhaps place the three plots with identical xaxis below each other. It would make it easy for the reader to get a quick overview.
- If there is no secondary yaxis, then always place the yaxis to the left.
- Be consistent about using the term observation and the actual name of the observation product in the legend

Response: Thanks for the suggestions. We believe your suggestions are very helpful to improve the readability of figure 6. In the revised manuscript, we will place the Y-axis to the left in subplot (b) and (d) in Figure 6 and we will change the legends in subplot (a), (c), and (d) to be "Gauge observation".

Page 13 line 270: *"The experiment of f_{drain}…"*. It would be easier to read and understand the text if the naming of the experiments were consistent. This apply to the whole paper.

Response: Thanks for the comment. In the revised manuscript, we will consistently use the naming of the experiments that we defined in Table 1 everywhere.

Figure 7: Please explain the term "ET trend" in the text. As I understand, you use the term "trend" regarding runoff in the paper. Please clarify in the text which trend you are referring to.

Response: In Figure 7, ET trend refers to the Sen's slope of annual basin averaged ET during the simulation period (i.e., 1979-2009). We will clarify it in the caption of Figure 7 in the revised manuscript.

Figure 8: There is something with the naming, why c and d?

- I suggest making the figure in the same way as figure 6, and with the same order of the plots.

Response: The figure subplot titles were messed up when we converted the figure format. We will correct it in the revised manuscript.

Figure 9: regarding legend in (a) and c): Be consistent with naming of the experiments in relation to text and other figures. The whole article would be much more readable if you were consistent with the naming.

Response: We appreciate this comment. We will modify the figure 9 in the revised manuscript to have consistent naming.

Page 16, line 320: There is no eq. 7

Response: Sorry for the confusion. It should be the objective function, i.e., Eq (4).

Figure 10: c): the dotted line is missing in legend

Response: The dotted line represents annual time series from observation. We didn't include it in the legend of Figure 10 because the Trend and p-value of the observation were reported in Figure 5. We found adding the observation in the legend will make Figure 10 busy. In the revised manuscript, we will describe the dotted line in the caption with relevant metrics.

Reference

Xu, D., Bisht, G., Sargsyan, K., Liao, C. and Leung, L.R. 2022. Using a surrogate-assisted Bayesian framework to calibrate the runoff-generation scheme in the Energy Exascale Earth System Model (E3SM) v1. Geosci. Model Dev. 15(12), 5021-5043.